

# **Local Weighting–Based Diagnostics for Bayesian Poststratification**

---

Ryan Giordano, Alice Cima, Erin Hartman, Jared Murray, Avi Feller  
**Berkeley BSTARS September 2025**

# Are US non-voters becoming more Republican?

## Blue Rose research says yes:

“Politically disengaged voters have become much more Republican, And because less-engaged voters swung away from [Democrats], an expanded electorate meant a more Republican electorate.”

[Blue Rose Research, 2024] (On Ezra Klein show, major professional pollsters)

## On Data and Democracy says no:

“Claims of a decisive pro-Republican shift among the overall non-voting population are not supported by the most reliable, large-scale post-election data currently available.”

[Bonica et al., 2025] (Berkeley professor co-author, major professional researchers)

- 
- The problem is very hard (it's difficult to accurately poll non-voters)
  - Different data sources
  - **Very different statistical methods:** ★
    - Blue Rose uses Bayesian hierarchical modeling (MrP)
    - The CES uses weighted averages (calibration weighting)

## Our contribution

We provide a calibration weighting interpretation of MrP analyses that:

- Is easily computable from MCMC draws and standard software, and
- Defines MrP versions of key diagnostics that motivate calibration weighting.

We provide apples-to-apples comparisons between MrP and calibration weighting.

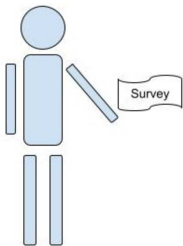
- Introduce the statistical problem and two methods (calibration weighting and MrP)
- Describe one of the classical calibration weighting diagnostics (covariate balance)
- Define MrPaw & state a key theorem
- Real-world results
- Future directions

# The basic problem

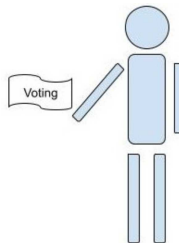
We have a survey population, for whom we observe:

- Covariates  $\mathbf{x}$  (e.g. race, gender, zip code, age, education level)
- Responses  $y$  (e.g. A binary response to “do you support policy such-and-such”)

We want the average response in a target population, in which we observe only covariates.



Observe  $(\mathbf{x}_s, y_s)$  for  $s = 1, \dots, S$



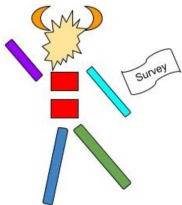
Observe  $\mathbf{x}_t$  for  $t = 1, \dots, T$

# The basic problem

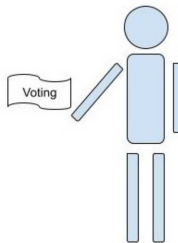
We have a survey population, for whom we observe:

- Covariates  $\mathbf{x}$  (e.g. race, gender, zip code, age, education level)
- Responses  $y$  (e.g. A binary response to “do you support policy such-and-such”)

We want the average response in a target population, in which we observe only covariates.



Observe  $(\mathbf{x}_s, y_s)$  for  $s = 1, \dots, S$



Observe  $\mathbf{x}_t$  for  $t = 1, \dots, T$

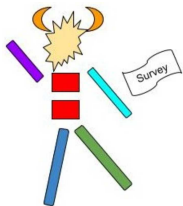
**The problem is that the populations are very different.**

# The basic problem

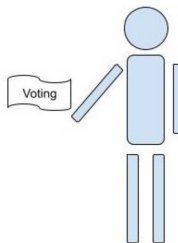
We have a survey population, for whom we observe:

- Covariates  $\mathbf{x}$  (e.g. race, gender, zip code, age, education level)
- Responses  $y$  (e.g. A binary response to “do you support policy such-and-such”)

We want the average response in a target population, in which we observe only covariates.



Observe  $(\mathbf{x}_s, y_s)$  for  $s = 1, \dots, S$



Observe  $\mathbf{x}_t$  for  $t = 1, \dots, T$

**The problem is that the populations are very different.**

Our survey results may be biased.

How can we use the covariates to say something about the target responses?

## Two approaches

We want  $\mu := \frac{1}{T} \sum_{t=1}^T y_t$ , but don't observe target population  $y_t$ .

- Assume  $p(y|\mathbf{x})$  is the same in both populations,
- But the distribution of  $\mathbf{x}$  may be different in the survey and target.

## Two approaches

We want  $\mu := \frac{1}{T} \sum_{t=1}^T y_t$ , but don't observe target population  $y_t$ .

- Assume  $p(y|\mathbf{x})$  is the same in both populations,
- But the distribution of  $\mathbf{x}$  may be different in the survey and target.

### Calibration weighting

- Choose “calibration weights”  $w_s$   
(e.g. raking weights)

### Bayesian hierarchical modeling (MrP)

- Choose model  $\mathcal{P}(y|x, \theta)$  and prior  $\mathcal{P}(\theta)$   
(e.g. Hierarchical logistic regression)



## Two approaches

We want  $\mu := \frac{1}{T} \sum_{t=1}^T y_t$ , but don't observe target population  $y_t$ .

- Assume  $p(y|\mathbf{x})$  is the same in both populations,
- But the distribution of  $\mathbf{x}$  may be different in the survey and target.

### Calibration weighting

- Choose “calibration weights”  $w_s$   
(e.g. raking weights)
- Take  $\hat{\mu}_{\text{CAL}} = \frac{1}{S} \sum_{s=1}^S w_s y_s$

### Bayesian hierarchical modeling (MrP)

- Choose model  $\mathcal{P}(y|x, \theta)$  and prior  $\mathcal{P}(\theta)$   
(e.g. Hierarchical logistic regression)
- Take  $\hat{y}_t = \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})} [y|\mathbf{x}_t]$  and  
 $\hat{\mu}_{\text{MRP}} = \frac{1}{T} \sum_{t=1}^T \hat{y}_t$

## Two approaches

We want  $\mu := \frac{1}{T} \sum_{t=1}^T y_t$ , but don't observe target population  $y_t$ .

- Assume  $p(y|\mathbf{x})$  is the same in both populations,
- But the distribution of  $\mathbf{x}$  may be different in the survey and target.

### Calibration weighting

- Choose “calibration weights”  $w_s$   
(e.g. raking weights)
- Take  $\hat{\mu}_{\text{CAL}} = \frac{1}{S} \sum_{s=1}^S w_s y_s$
- Dependence on  $y_s$  is obvious  
( $w_s$  typically chosen using only  $\mathbf{x}$ )

### Bayesian hierarchical modeling (MrP)

- Choose model  $\mathcal{P}(y|x, \theta)$  and prior  $\mathcal{P}(\theta)$   
(e.g. Hierarchical logistic regression)
- Take  $\hat{y}_t = \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})} [y|\mathbf{x}_t]$  and  
 $\hat{\mu}_{\text{MRP}} = \frac{1}{T} \sum_{t=1}^T \hat{y}_t$
- Dependence on  $y_s$  very complicated  
(Typically via MCMC draws from  $\mathcal{P}(\theta|\text{Survey data})$ )

## Two approaches

We want  $\mu := \frac{1}{T} \sum_{t=1}^T y_t$ , but don't observe target population  $y_t$ .

- Assume  $p(y|\mathbf{x})$  is the same in both populations,
- But the distribution of  $\mathbf{x}$  may be different in the survey and target.

### Calibration weighting

- Choose “calibration weights”  $w_s$   
(e.g. raking weights)

- Take  $\hat{\mu}_{\text{CAL}} = \frac{1}{S} \sum_{s=1}^S w_s y_s$

- Dependence on  $y_s$  is obvious  
( $w_s$  typically chosen using only  $\mathbf{x}$ )

- Weights give interpretable diagnostics:

- Frequentist variability
- Partial pooling
- Regressor balance

### Bayesian hierarchical modeling (MrP)

- Choose model  $\mathcal{P}(y|x, \theta)$  and prior  $\mathcal{P}(\theta)$   
(e.g. Hierarchical logistic regression)

- Take  $\hat{y}_t = \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})} [y|\mathbf{x}_t]$  and  
 $\hat{\mu}_{\text{MRP}} = \frac{1}{T} \sum_{t=1}^T \hat{y}_t$

- Dependence on  $y_s$  very complicated  
(Typically via MCMC draws from  
 $\mathcal{P}(\theta|\text{Survey data})$ )

- **Black box**

## Two approaches

We want  $\mu := \frac{1}{T} \sum_{t=1}^T y_t$ , but don't observe target population  $y_t$ .

- Assume  $p(y|\mathbf{x})$  is the same in both populations,
- But the distribution of  $\mathbf{x}$  may be different in the survey and target.

### Calibration weighting

- ▶ Choose “calibration weights”  $w_s$   
(e.g. raking weights)
- ▶ Take  $\hat{\mu}_{\text{CAL}} = \frac{1}{S} \sum_{s=1}^S w_s y_s$
- ▶ Dependence on  $y_s$  is obvious  
( $w_s$  typically chosen using only  $\mathbf{x}$ )
- ▶ Weights give interpretable diagnostics:
  - Frequentist variability
  - Partial pooling
  - Regressor balance

### Bayesian hierarchical modeling (MrP)

- ▶ Choose model  $\mathcal{P}(y|x, \theta)$  and prior  $\mathcal{P}(\theta)$   
(e.g. Hierarchical logistic regression)
- ▶ Take  $\hat{y}_t = \mathbb{E}_{\mathcal{P}(\theta|\text{Survey data})} [y|\mathbf{x}_t]$  and  
 $\hat{\mu}_{\text{MRP}} = \frac{1}{T} \sum_{t=1}^T \hat{y}_t$
- ▶ Dependence on  $y_s$  very complicated  
(Typically via MCMC draws from  
 $\mathcal{P}(\theta|\text{Survey data})$ )
- ▶ **Black box**  
← (We open this box, providing analogues  
of all these diagnostics)

# What are we weighting for?<sup>1</sup>

We want:

$$\text{Target average response} = \frac{1}{T} \sum_{t=1}^T y_p \approx \frac{1}{S} \sum_{s=1}^S w_s y_s = \text{Weighted survey average response}$$

We can't check this, because we don't observe  $y_p$ .

---

<sup>1</sup>Pun attributable to Solon et al. [2015]

# What are we weighting for?<sup>1</sup>

We want:

$$\text{Target average response} = \frac{1}{T} \sum_{t=1}^T y_p \approx \frac{1}{S} \sum_{s=1}^S w_s y_s = \text{Weighted survey average response}$$

We can't check this, because we don't observe  $y_p$ . But we can check whether:

$$\frac{1}{T} \sum_{t=1}^T \mathbf{x}_p = \frac{1}{S} \sum_{s=1}^S w_s \mathbf{x}_s$$

Such weights satisfy “covariate balance” for  $\mathbf{x}$ .

You can check covariate balance for any calibration weighting estimator.

---

<sup>1</sup>Pun attributable to Solon et al. [2015]

# What are we weighting for?<sup>1</sup>

We want:

$$\text{Target average response} = \frac{1}{T} \sum_{t=1}^T y_p \approx \frac{1}{S} \sum_{s=1}^S w_s y_s = \text{Weighted survey average response}$$

We can't check this, because we don't observe  $y_p$ . But we can check whether:

$$\frac{1}{T} \sum_{t=1}^T \mathbf{x}_p = \frac{1}{S} \sum_{s=1}^S w_s \mathbf{x}_s$$

Such weights satisfy “covariate balance” for  $\mathbf{x}$ .

You can check covariate balance for any calibration weighting estimator.

Even more, covariate balance is the criterion for a popular class of calibration weight estimators:

## Raking calibration weights

“Raking” selects weights that

- Are as “close as possible” to some reference weights
- Under the constraint that they balance some selected regressors.

---

<sup>1</sup>Pun attributable to Solon et al. [2015]

## Generalized covariate balance checks

We want to balance  $f(\mathbf{x})$  because we think  $\mathbb{E}[y|\mathbf{x}]$  might plausibly vary  $\propto f(\mathbf{x})$ , and want to check whether our estimator can capture this variability.

This motivates the following **generalized covariate balance check**:



# Generalized covariate balance checks

We want to balance  $f(\mathbf{x})$  because we think  $\mathbb{E}[y|\mathbf{x}]$  might plausibly vary  $\propto f(\mathbf{x})$ , and want to check whether our estimator can capture this variability.

This motivates the following **generalized covariate balance check**:

## Generalized covariate balance (GCB) (informal)

Pick a small  $\delta$ , and define a *new response variable*  $\tilde{y}$  such that

$$\mathbb{E}[\tilde{y}|\mathbf{x}] = \mathbb{E}[y|\mathbf{x}] + \delta f(\mathbf{x}).$$

We know the change this is supposed to induce in the target population.

Covariate balance checks whether our estimators produce the same change.

# Generalized covariate balance checks

We want to balance  $f(\mathbf{x})$  because we think  $\mathbb{E}[y|\mathbf{x}]$  might plausibly vary  $\propto f(\mathbf{x})$ , and want to check whether our estimator can capture this variability.

This motivates the following **generalized covariate balance check**:

## Generalized covariate balance (GCB) (formal)

Pick a small  $\delta$ , and define a *new response variable*  $\tilde{y}$  such that

$$\mathbb{E}[\tilde{y}|\mathbf{x}] = \mathbb{E}[y|\mathbf{x}] + \delta f(\mathbf{x}).$$

We know the expected change this perturbation produces in the target distribution:

$$\mathbb{E}[\mu(\tilde{y}) - \mu(y)|\mathbf{x}] = \frac{1}{T} \sum_{t=1}^T (\mathbb{E}[\tilde{y}|\mathbf{x}_p] - \mathbb{E}[y|\mathbf{x}_p]) = \delta \frac{1}{T} \sum_{t=1}^T f(\mathbf{x}_p)$$

Then, check whether your estimator  $\hat{\mu}(\cdot)$  produces the same change:

$$\underbrace{\hat{\mu}(\tilde{y}) - \hat{\mu}(y)}_{\substack{\text{Replace weighted averages} \\ \text{with changes in an estimator}}} \stackrel{\text{check}}{=} \delta \frac{1}{T} \sum_{t=1}^T f(\mathbf{x}_p).$$

# Generalized covariate balance checks

We want to balance  $f(\mathbf{x})$  because we think  $\mathbb{E}[y|\mathbf{x}]$  might plausibly vary  $\propto f(\mathbf{x})$ , and want to check whether our estimator can capture this variability.

This motivates the following **generalized covariate balance check**:

## Generalized covariate balance (GCB) (formal)

Pick a small  $\delta$ , and define a *new response variable*  $\tilde{y}$  such that

$$\mathbb{E}[\tilde{y}|\mathbf{x}] = \mathbb{E}[y|\mathbf{x}] + \delta f(\mathbf{x}).$$

We know the expected change this perturbation produces in the target distribution:

$$\mathbb{E}[\mu(\tilde{y}) - \mu(y)|\mathbf{x}] = \frac{1}{T} \sum_{t=1}^T (\mathbb{E}[\tilde{y}|\mathbf{x}_p] - \mathbb{E}[y|\mathbf{x}_p]) = \delta \frac{1}{T} \sum_{t=1}^T f(\mathbf{x}_p)$$

Then, check whether your estimator  $\hat{\mu}(\cdot)$  produces the same change:

$$\underbrace{\hat{\mu}(\tilde{y}) - \hat{\mu}(y)}_{\substack{\text{Replace weighted averages} \\ \text{with changes in an estimator}}} \stackrel{\text{check}}{=} \delta \frac{1}{T} \sum_{t=1}^T f(\mathbf{x}_p).$$

When  $\hat{\mu}(\cdot) = \hat{\mu}_{\text{CAL}}(\cdot)$ , GCB recovers the standard covariate balance check.

**Step one:** Construct  $\tilde{y}$  such that  $\mathbb{E} [\tilde{y}|\mathbf{x}] = \mathbb{E} [y|\mathbf{x}] + \delta f(\mathbf{x})$ .

## Generalized covariate balance for MrP

**Step one:** Construct  $\tilde{y}$  such that  $\mathbb{E} [\tilde{y}|\mathbf{x}] = \mathbb{E} [y|\mathbf{x}] + \delta f(\mathbf{x})$ .

**Problem:** Our  $y$  is binary! (We're motivated by hierarchical linear regression.)

# Generalized covariate balance for MrP

**Step one:** Construct  $\tilde{y}$  such that  $\mathbb{E} [\tilde{y}|\mathbf{x}] = \mathbb{E} [y|\mathbf{x}] + \delta f(\mathbf{x})$ .

**Problem:** Our  $y$  is binary! (We're motivated by hierarchical linear regression.)

Two possibilities:

- Allow  $\tilde{y}$  to take values other than  $\{0, 1\}$  and set  $\tilde{y} = y + \delta f(\mathbf{x})$ , or
- Use an estimate of  $\mathbb{E} [y|\mathbf{x}]$  to draw new binary  $\tilde{y}$ .

We define theory and methods for the first, and provide tools for generating data using the second method for potentially problematic regressors.

## Generalized covariate balance for MrP

**Step one:** Take  $\tilde{y} = y + \delta f(\mathbf{x})$ .

**Step two:** Evaluate  $\hat{\mu}_{\text{MRP}}(\tilde{y}) - \hat{\mu}(y)$ .

## Generalized covariate balance for MrP

**Step one:** Take  $\tilde{y} = y + \delta f(\mathbf{x})$ .

**Step two:** Evaluate  $\hat{\mu}_{\text{MRP}}(\tilde{y}) - \hat{\mu}(y)$ .

**Problem:**  $\hat{\mu}_{\text{MRP}}(\cdot)$  is computed with MCMC.

- Takes hours to re-run, and
- Output is noisy, and  $\hat{\mu}_{\text{MRP}}(\tilde{y}) - \hat{\mu}(y)$  may be small.



# Generalized covariate balance for MrP

**Step one:** Take  $\tilde{y} = y + \delta f(\mathbf{x})$ .

**Step two:** Evaluate  $\hat{\mu}_{\text{MRP}}(\tilde{y}) - \hat{\mu}(y)$ .

**Problem:**  $\hat{\mu}_{\text{MRP}}(\cdot)$  is computed with MCMC.

- Takes hours to re-run, and
- Output is noisy, and  $\hat{\mu}_{\text{MRP}}(\tilde{y}) - \hat{\mu}(y)$  may be small.

## Taylor series

Form the approximation

$$\hat{\mu}_{\text{MRP}}(\tilde{y}) = \sum_{s=1}^S w_s^{\text{MRP}} (\tilde{y}_s - y_s) + \text{Residual} \quad \text{where} \quad w_s^{\text{MRP}} := \frac{d}{dy_s} \hat{\mu}_{\text{MRP}}(y).$$

If MrP were linear (e.g. if you use OLS instead of hierarchical logistic regression), then

- The residual is zero,
- $\hat{\mu}_{\text{MRP}}(y) = \sum_{s=1}^S w_s^{\text{MRP}} y_s$ , and so
- $\hat{\mu}_{\text{MRP}}(\tilde{y})$  is a calibration weighting estimator, and  $w_s^{\text{MRP}}$  are its weights. (Cite Gelman)

In general, MrP is truly nonlinear. The residual is only small when  $\tilde{y} \approx y$  (i.e., when  $\delta \ll 1$ ).

# Generalized covariate balance for MrP

**Step one:** Take  $\tilde{y} = y + \delta f(\mathbf{x})$ .

**Step two:** Evaluate  $\hat{\mu}_{\text{MRP}}(\tilde{y}) - \hat{\mu}(y)$ .

**Problem:**  $\hat{\mu}_{\text{MRP}}(\cdot)$  is computed with MCMC.

- Takes hours to re-run, and
- Output is noisy, and  $\hat{\mu}_{\text{MRP}}(\tilde{y}) - \hat{\mu}(y)$  may be small.

## Taylor series

Form the approximation

$$\hat{\mu}_{\text{MRP}}(\tilde{y}) = \sum_{s=1}^S w_s^{\text{MRP}} (\tilde{y}_s - y_s) + \text{Residual} \quad \text{where} \quad w_s^{\text{MRP}} := \frac{d}{dy_s} \hat{\mu}_{\text{MRP}}(y).$$

It happens that the needed derivatives are given by simple a posteriori covariances involving only the inverse link function  $m(\mathbf{x}; \theta)$  and log likelihood [Giordano et al., 2018].

These can be computed using standard MCMC software (e.g. Bürkner [2017]).

## Theorem

- Let  $\tilde{y} = y + \delta f(\mathbf{x})$ ,
- $\hat{\mu}_{\text{MRP}}$  be a hierarchical logistic regression posterior expectation, and
- $\mathcal{F}$  be a Donsker class of uniformly bounded functions on  $\mathbf{x}$ .

Then, with probability approaching one, as  $N \rightarrow \infty$ ,

$$\sup_{f \in \mathcal{F}} \left( \hat{\mu}_{\text{MRP}}(\tilde{y}) - \left( \hat{\mu}_{\text{MRP}}(y) + \sum_{s=1}^S w_s^{\text{MRP}} \delta f(\mathbf{x}_s) \right) \right) = O(\delta^2) \quad \text{as } \delta \rightarrow 0$$

The supremum over  $\mathcal{F}$  is the primary technical contribution! It means we are justified in searching over regressors to find imbalance.

Draws on our prior work on uniform and finite-sample error bounds for Bernstein–von Mises theorem–like results [Giordano and Broderick, 2024, Kasprzak et al., 2025].

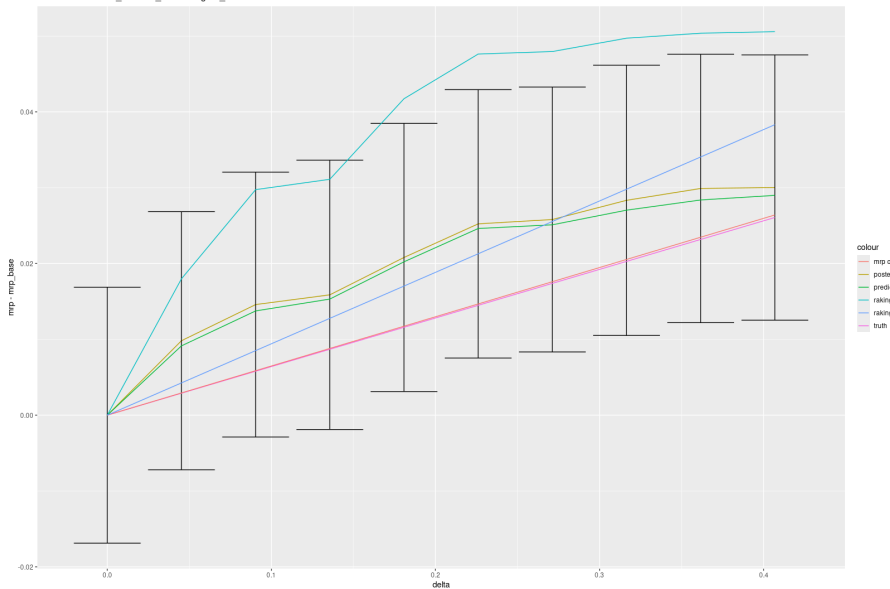
Show using the Alexander experiment that

- The MrPaw approximation is accurate, even up to a standard deviation away
- Differential imbalance actually exists between raking and MrP and we can detect it
- We can make binary datasets for the changes, and there's variability in how to do that

Remind that we can do frequentist variance and partial pooling too.

# Experiment

alexander decade\_married\_rk1989region\_rkwest



- Instance of a very general class of local consistency checks that generalize classical regression checks (work with Sequoia)
- Versions for GLMMs (work with Vladimir)
- Going beyond classical Bayesian sensitivity (work with Lucas)

- Blue Rose Research. 2024 Election Retrospective Presentation.  
<https://data.blueroseresearch.org/2024retro-download>, 2024. Accessed on 2024-10-26.
- A. Bonica, R. Fordham, J. Grumbach, and E. Tiburcio. Did non-voters really flip Republican in 2024? The evidence says no.  
<https://data4democracy.substack.com/p/did-non-voters-really-flip-republican>, April 2025.
- Paul-Christian Bürkner. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1): 1–28, 2017. doi: 10.18637/jss.v080.i01.
- R. Giordano and T. Broderick. The Bayesian infinitesimal jackknife for variance, 2024. URL <https://arxiv.org/abs/2305.06466>.
- R. Giordano, T. Broderick, and M. I. Jordan. Covariances, robustness and variational bayes. *Journal of machine learning research*, 19(51), 2018.
- M. Kasprzak, R. Giordano, and T. Broderick. How good is your Laplace approximation of the bayesian posterior? Finite-sample computable error bounds for a variety of useful divergences, 2025. URL <https://arxiv.org/abs/2209.14992>.
- G. Solon, S. Haider, and J. Wooldridge. What are we weighting for? *Journal of Human resources*, 50(2):301–316, 2015.